Sentiment Classification of Social Media Data with Supervised Machine Learning Approaches: Common Framework, Challenges, and New Dimensions

Suresha N. Perera[1], Kasun Karunanayaka[2]
[1]Department of Computer Science, Open University of Sri Lanka, Nawala, Sri Lanka
[2]University of Colombo School of Computing, Reid Avenue, Colombo 7, Sri Lanka
Email: [1]waper@ou.ac.lk, [2]ktk@ucsc.cmb.ac.lk

***Abstract*—As an interdisciplinary research field, sentiment analysis is one of the momentous applications in Natural Language Processing, for quantifying the emotional value of text available in social media networks. The mission of this paper is to present a comprehensive overview of recent Machine Learning sentiment analysis classification techniques (Naive Bayes, Support Vector Machine, Rough-Fuzzy based classifiers etc.) to serve the scholars and researchers by emphasizing the methods used in current research. However, there is a minimal number of review papers discussed rough-fuzzy classifier involvement by researchers in sentiment analysis and there is a plethora of work that must be done with text mining. In addition to those, we propose a common framework for sentiment analysis in the context of social media based on previous works by providing the facility for users to enhance it with new concepts. Finally, discuss various research challenges and possible future directions in sentiment analysis.**

Keywords—Sentiment Analysis, Rough-Fuzzy Classifiers, Natural Language Processing, Machine Learning, Social Networking Sites, Introduction

I. INTRODUCTION

Decision-making behavior in an online community is predominant because many individuals now use community-based web services thus the opinions of others are readily available in online environments where people often rely on other individuals' decisions for social validation to make their own. For instance, individuals are inter-ested in others' opinions about political candidates before making a voting decision in a political election. In order to gather the public and consumer opinion organizations conduct surveys, opinion polls, or focus groups. The power of Social Media Websites (SMW) is rampant and growing a plethora of information and data convoluted with varying interests, opinions, and emotions with human generated baselines. The wide-spread use of the above media encourages positive and negative attitudes about people, organizations, places, events, and ideas. The growth of social media usage has led to an increasing accumulation of data and it is used for the prediction of the outcome as this platform becomes the most powerful communication media on the internet while hundreds of millions of messages are being posted every day.

SA, also called opinion mining, which is the field of computational study that analyzes people's opinions, attitudes and emotions toward an entity. It repre-sents a large problem space with many names and slightly different tasks, e.g., SA, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc. There are three different levels of SA that has been investigated as:

Document Level Analysis - Classify whether a whole opinion document ex-presses a positive, negative or neutral sentiment [1], [2].

Sentence Level Analysis - Polarity is calculated for each while determining whether each sentence expressed a positive, negative, or neutral opinion.

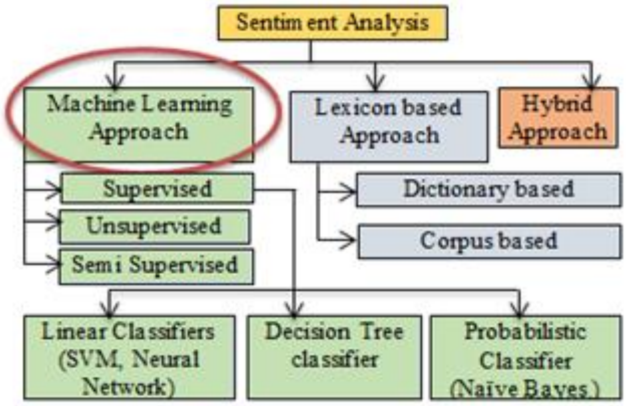

Fig 1 Three main approaches of the Sentiment Classification Method and it is further classified into subcategories. Only the ML branch is considered in this study.

Entity/Aspect Level Analysis - This discovers sentiments on entities and/or their aspects.

The three main approaches addressing SA [3] with social media are Lexicon based approach, ML based approach and Hybrid based approach [4] as shown in Fig. 1.

Lexicon based approach utilizes a sentiment lexicon to describe the polarity as positive, negative or neutral of a textual content and need the involvement of a human being at the analysis phase. This approach can be further divided into two categories: Dictionary based approach and Corpus based approach. ML approach can be divided into supervised, unsupervised and semi-supervised methods; it requires a large data set to be effective which is the main drawback. The classifier is trained on labeled data similar to test data with supervised learning. With unsupervised classifications, it is assigned labels based only on internal differences between the data points. ML provides more accuracy than lexicon based approach. Many researchers focused on Naive Bayes (NB) classification and Support Vector Machine (SVM) under the ML approach. Hybrid approach is the amalgamation of both ML and lexicon-based methods where many scholars are focusing nowadays. Supervised ML techniques have shown relatively better performance than the unsupervised lexicon based methods [5]. Most of the algorithms used in ML approach belong to supervised classification and provide higher accuracy and performance [6], [7].

This paper presents a comprehensive literature review on different text mining and SA approaches with ML techniques in distinct areas by identifying future directions on research. A common framework will be discussed in this research paper for the use of researches to start SA with social media data. There are many challenges with SA applications which direct for new opportunities. This becomes more challenging as this is to determine the emotional state of a person with data mining concepts. Finally, the difficulties and challenges of text SA are discussed. SA on SMW data has become increasingly popular among the academics and they have undertaken a diverse range of related research.

This paper is articulated as follows: Section II presents related work. Section III describes a classification framework with supervised ML techniques. Section IV discusses challenges arising of ML techniques for SA. Finally, Section V leads to the conclusion.

II. LITERATURE REVIEW

There exists substantial research on SA and most active research on the area came with the explosion of user-generated content in social media, discussion forums, blogs and reviews. Since most studies use or depend on ML approaches, the amount of user generated content provided unlimited data for training. This literature review provides an overview of the different approaches that can be applied for SA as well as brief explanations of algorithms used by researchers.

To predict heart disease, Srinivas et al. [8] have exploited a rough-fuzzy classifier. It combines rough set with fuzzy set theory. The rules were generated using rough set theory and prediction done using a fuzzy classifier. The rule generation and relevant attribute identification are automated with rough sets and rules are automatically applied to fuzzy classifier for prediction of heart disease. The above classifier can be extended by using the associative analysis to find the relevant attribute and incorporating statistical measures to strengthen the computation. Keerthika et al [9] construct a Temporal Fuzzy Rule Based Classifier (TFRBC) which is built by the generalization of Fuzzy Rough Sets and uses fuzzy rough set. Further they have used temporal logic for mining temporal patterns in medical databases. After comparing the accuracies of classifiers such as FNN (Fuzzy Neural Network) and TFRBC it can be concluded that the accuracy of TFRBC is more with 88% when compared with the FNN which is 74%. As a future direction, the rule based classifier can be improved with effective decisions by taking upper approximations while reducing the number of rule sets when compared with the lower approximations.

Srividya and Sowjanya [10] discussed a methodology to analyse collected reviews over a period of time in Facebook using NB classifier into relevant and non-relevant to determine public opinion on the popularity of android based and identified the most preferable versions of OS for android phones and found Android KitKat is more preferable than others.

Election forecasting has a political side which helps in forecasting the results. In order to forecast an election there are opinion polls as well as used and many scientifically proven statistical models [11]. However sometimes polls also fail in predicting the results of the election even in the developed countries. According to [12], it listed several failed polls results such as in the 1992 British General Elections, French presidential etc.

Electoral analysis of twitter data is straightforward and optimistic even though it is a challenge for the research community in today's world. The outcome of the election with twitter data was analyzed by Salunkhe and Deshmukh [13] with the US and Gujarat Rajya Sabha election. The data collection was based on two methods: keyword match with tweets and using twitter API the collected data were pre-proceed with lower case conversion, punctuation and number removal, stemming and striping white spaces. NB used on the above training data set including emotions for the sentimental analysis. The dictionary based approach with eleven lexicons used for the classification of identified variables. The case study analysis concludes that SMW like twitter can be used in elections to predict the outcome well in advance. Further, this helps to explore the sentiment or views of citizens who have the voting power. Moreover, it can be successfully influenced by voters.

Wongn et al. [14] used twitter streaming API to collect tweets which contain the identified keyword relevant to the events identified before the election and used lexicon-based SA package, to extract the sentiment of tweets as a ternary (positive, negative, neutral) classification. They used the consistency relationship between tweeting and retweeting behavior. It is evident that the proposed method operates at much faster than existing methods. However, it does not require the explicit knowledge of the twitter network. As future directions they proposed to get the use of retweet matrix and retweet average scores when developing new models and algorithms.

Vadivukarassi et al. [15] proposed a model to analyse twitter data and preprocessed using Natural Language Toolkit (NTLK) techniques. The word scores of the features are tested based on Chi-square method and key words were scored sentiments during the analysis of data. NB classifier is used for training and testing the features and also evaluated the sentimental polarity and generated summarized report about the opinion from twitter.

Prabhu et al. [16] proposed a methodology to distribute political party’s tickets during the election with twitter data. Upon receiving all the necessary data related to a candidate, NB Algorithm used to predict the most deserving candidate. Their final conclusion was that this classifier is suitable for any type of election in India and elsewhere.

Pak and Paroubek [17] proposed a classifier with collected twitter corpus to conduct linguistic analysis. This classifier tested only for the English language and the proposed technique can be used with any other language as future work. For the analysis of the corpus, they used a plot of word frequencies with Zipf’s law inorder to understand how terms are distributed across collected corpus and TreeTagger [18] which is a tool for annotating text with part-of-speech and lemma information. Under the training of the classifier they followed the method of filtering, tokenization, removal of stop words and constructing n-grams from twitter posts. As the classifier NB was selected and it outperformed with best results. The accuracy was established by discarding common n-grams with two strategies: computing the entropy of a probability distribution of the appearance of an n-gram in different datasets [19] and introducing a new term "salience" calculated for each n-gram and used multinomial NB.

Rao et al. [20] identified the importance of identifying the general sentiment polarity of a news article before publishing. They use a ML approach on twitter data with different features like unigram, bigram and hybrid. It shows that the hybrid feature with SVM classifier gives the best results for prediction of sentiment of twitter data. This Classifier can be further developed to make automatic sentiment classifiers for more than one language.

Wei and Gulla [21] present an analysis technique based on a tree of feelings of ontological features. The product's attributes labeling is handled with the novel HL-SOT approach. Hierarchical Learning (HL) process used for analyzing their associated sentiments in product reviews. Further they used a defined Sentiment Ontology Tree (SOT) with the above process. The HL-flat approach ignores the hierarchical relationships among labels when training each classifier and this is a "flat" version of HL-SOT. The H-RLS algorithm only uses identical threshold values for each classifier in the classification process where HL-SOT enables the threshold values to be learned separately for each classifier in the training process. The research found that the HL-SOT approach outperforms two baselines: the HL-flat and the H-RLS approach. The classification approach used is based on hierarchical classification algorithms.

Pang and Lee [2], [22] conducted a study to compare the performance of NB, Maximum Entropy and SVM for SA based on different features. For instance they considered only unigrams, bigrams, combination of both, incorporating parts of speech and position information, taking only adjectives etc. It is observed from the results that the feature presence is more important than feature frequency. Further, the accuracy falls when using bigrams. Moreover the accuracy improved when considering all the frequently occurring words from all parts of speech. Furthermore, the accuracy improved with not only with

adjectives but also with including position information. It was evident that NB performs better than SVM when the feature space is small, but SVM's perform better when feature space is increased.

A study has been conducted [23] on Pang Corpus which is a movie review database and opinions collected from the website Epinions.com named as Taboada Corpus to train a sentiment classifier with SVM together with n-grams and different weighting schemes: Term Frequency Inverse Document Frequency (TFIDF), Binary Occurrence (BO) and Term Occurrence (TO). Further, it uses chi-square weight features to select informative features. It is evident that the chi-square feature selection improves the accuracy of the classification. Further, it shows that the unigrams outperform the other n-gram models for both datasets. Table 1 depcted the summary of some of the researches conducted by scholars with their used algorithms, outcomes, advantages, dis-advantages and futre works.

Table 1. Sentiment analysis of social media data with various supervised learning techniques with their used algorithms, outcomes, advantages, dis-advantages, and future works

Study	**Dataset**	**classifier/ Algorithm**	**Outcomes/Advantages/ Disadvantages/Future works**
[8]	Cleveland, Hungarian and Switzerland datasets for heart disease prediction	Rough-fuzzy classifier	- Outperformed the previous approaches **Future work:** - Can be extended by including the associative analysis to find the relevant attribute

			- Rule strength computation can be extended by including statistical measure
[9]	Diabetic dataset to mine temporal patterns in medical databases	Fuzzy Temporal Rule Based Classifier	**Advantages:** - The minimum number of human involvement - Classifier is simple **Future work** - Build the rule based classifier with upper approximations with effective decisions to reduce the number of rule sets - Use less number of rule sets in order to build an efficient rule based classifier
[13]	Tweets in US and Gujarat Rajya Sabha election	NB with dictionary based approach	- Make prediction of future outcome of the election - Extract the sentiment or views of people - Sentiment analysis to classify their sentiment.

[15]	Tweets in 2012 U.S. presidential election campaign	Chi-Square test and NB classifier	- When the number of features increases, the accuracy of the selected features also increases. - Easy to generate summary report about the opinion from Twitter
[17]	Real Twitter posts	Multinomial NB classifier that uses N-gram and POS-tags as features	**Advantages:** - Improve the performance of the system by increasing the sample size - Best performance is achieved when using bigrams - When filtering out the common n-grams: salience provides a better accuracy than the entropy - Attaching negation words when forming n-grams increase performance **Future work:**

			use a multilingual corpus of twitter data and compare the characteristics of the corpus across different languages.
[20]	Extract twitter data using twitter API with analysis of news data	NB, SVM and maximum entropy using unigram, bigram and hybrid feature	- SVM using hybrid features outperforms the selection feature with accuracy of 84%. **Future work:** - Develop automatic sentiment classifiers for more than one language starting from the Hindi language.
[21]	Sentiment analysis on reviews of one product on customer reviews on digital cameras that are collected from a customer review website	A Hierarchical Learning (HL) process with a defined Sentiment Ontology Tree (SOT)	- HL-SOT approach outperforms two baselines: the HL-flat and the H-RLS approach **Future work:** - HLSOT approach can be easily generalized to labeling a mix of reviews of more than one products Advantages:

			- Separately learning threshold values for each classifier improve the classification accuracy
			- Knowledge of hierarchical relationships of labels improve the approach's performance
			- Product's attributes could be a useful knowledge for mining product review texts
			- lassification performance will be affected by variances of the generated SOTs
			an automatic method to learn a product's attributes and the structure of SOT from existing product review texts improve the efficiency

[23]	Pang Corpus and Taboada Corpus	SVM, N-grams and different weighting scheme, Chi-Square weight features to select informative features	- ChiSquare feature selection provide significant improvement on classification accuracy - Unigrams outperform other n-grams models - Binary Occurences (BO) and TFIDF weighting scheme plays a crucial role in extracting the most classical features in the data set - Accuracy is higher when the number of features selected is fewer - Performance varies with the domains and corpus size
[27]	Political tweets in presidential elections in Egypt 2012 for Arabic text classification	SVM and NB, with TF-IDF	- Obtained higher accuracy and performance with NB than SVM **Future work-** - Use Khoja stemmer and compare the result

			- Compare the result between unigram, bigram and trigram.
[30]	Twitter data	Competitive mod el that compares Linear (SVM) and prob abilistic approac h (Logistic Regression and NB)	- Observed that the performance and the accuracy is high with SVM **Future works** - Need more efficient machine learning, deep learning algorithm for better classifiers. - Deal with spam post/tweets.. - Use better mining techniques to deal with natural language processing more efficiently. **Findings:** **Advantages of SVM:** - Effective in high dimensional spaces.

			- Effective in cases where number of dimensions is greater than the number of samples. - Memory efficient. - Versatile **Disadvantages of SVM:** - Poor performances when the number of features is much greater than the number of samples. - Does not directly provide probability estimates and use an expensive five-fold cross-validation. **Advantages of NB:** - Easy and fast to predict class of test data set. - Performs well in multi class prediction. - Need less training data. - Performs well with categorical input variables compared to numerical variable(s).

			Disadvantages of NB: - Zero Frequency problem. Can use smoothing technique like Laplace estimation to solve this - The assumption of independent predictors.
[31]	Movie reviews from www.imdb.com and hotel reviews from OpinRank Review Dataset k+Review+Dataset)	K-Nearest Neighbour and NB algorithms	- Observed that the NB approach giving above 80% accuracies and outperforming than the k-NN approach. - Accuracies are lower for hotel reviews, in that both the classifiers yielded similar results. - NB classifier can be used successfully to analyse movie reviews **Future work -** - Compre with efficient sentiment analyser like random forest, Support vector Machine etc.

			- Implement a new algorithm utilizing the benefits of the both algorithms
[32]	Drug review analysis is collected by scraping from the raw HTML files using the Beautiful Soup Library in Python from Druglib.com	Use of fuzzy rough feature selection	- Fuzzy-rough feature selection significantly reduce the complexity of feature space - Reduce the classification run-time overheads while maintaining classification accuracy - Proposed framework bring forward more significant cost-efficiency savings to real-world healthcare analysis on large scale data Future work- - Exploit search strategies to increase the overall performance

			- Investigate the use of alternative approaches for learning classifiers which work better while dealing with the uncertainty inherent in natural language processing
[41]	The hashtagged (HASH) and emoticon (EMOT) as training datasets.	AdaBoost classifier,	- An F-measure of 0.68 was achieved for HASH. In addition, an F-measure of 0.65 was obtained by AdaBoost for HASH and EMOT datasets with a combination of n-grams, lexicons and microblogging features
		Unigrams, bigrams, lexicon, POS features, and micro-blogging features	- part-of-speech features may not be useful for sentiment analysis in the microblogging domain

			- An existing sentiment lexicon features were somewhat useful in conjunction with microblogging features. However, the microblogging features are the most useful. - hashtags are useful in collecting the data set.
[43]	Labeled movie dataset	NB and SVM algorithms Use CountVectorizer and TF-IDF	- SVM classifier outperforms every other classifier - Compare results of SVM with other supervised learning algorithms such as maximum entropy classifier, Stochastic gradient classifier, K nearest neighbor and others.

An optimized classifier proposed by Bhumika and Vaghelawith [24] evident that SVM outperformed existing systems. The study was conducted for the movie review, twitter and Gold dataset between Optimized SVM towards SVM and NB classifier. Modifying hyper parameter values of RBF kernel SVM gives better results compared to SVM and NB algorithms. The SVM was implemented with RBF kernel hyper parameter (C, γ) where those parameters are modified with different combinations of regularization Constant (Soft Margin) C , kernel hyperparameter γ (gamma). The proposed approach has found optimal value for hyper parameters obtaining more accuracy. In [25], a SA framework used to analyze the performance of SVM for textual polarity detection with pre-labeled two twitter and one IMDB reviews data sets adhering to four main steps as: data set, preprocessing, classification and results. For performance evaluation of SVM, three ratios of training data and test data are used: 70:30, 50:50 and 30:70. Research concluded that performance of SVM depends upon the dataset as well as on the ratio of Training and Test Data.

A recent research [1] has been conducted to analyze amazon product review dataset obtained from the UCI repository with the probabilistic classifier NB. The research has concluded that the NB classifier is a highly scalable, simple classifier technique which provides better level of accuracy and good results after classification. Further, it identified this technique can be applied to any kind of review dataset.

A case study conducted to carry qualitative analysis on SMW data related to political leaders to identify different sentiments [26]. The implemented multi-label classification algorithm is capable of classifying polarity. The results show that Tuning Multinomial NB performs better than NB. The survey revealed social media data on political domain which overcomes the major drawback of both manual qualitative analysis and large scale computational analysis of user generated textual content. Another study of twitter data [27] which focuses on presidential elections in Egypt 2012 was conducted and revealed that NB scores the highest accuracy and the lowest error rate for Arabic text classification. It opens new research areas: compare the results with other classifiers, use Khoja stemmer and compare it with light stemmer and compare the results with bigram and trigram. Ringsquandl and Petkovic [28] conducted a study to analyze

the presidential candidates of the Republican Party in the USA and their campaign topics. It found out that the special considerations in retrieval and pre-processing are needed, NLTK's built-in pre-processing functionalities were not sufficient for informal text corpora. As future work they identified the need of learning other domain-specific opinion words like nouns and verbs. Kassraie et al. [29] conducted a research on predicting the US 2016 elections results and found that not all of the voters are twitter and google users, social media isn't always reliable, having active spammer robots, etc. In order to control this as a future plan, the user behavior was tracked over time for validating the consistency or trend of their opinion.

Raghuwanshi and Pawar [30] performed a comparative study on NB, SVM, and Logistic regression with crowd source information that compares linear and probabilistic approach. The results revealed that SVM turns out to be best among all and can work with linear or non-linear data. Processing and extracting exact emotions are two major areas to work in this field, need more efficient machine learning, deep learning algorithms for better classifiers, a lot of ways to deal with spam posts/tweets and use better mining techniques to deal with NLP more efficiently.

A web crawling framework to facilitate the quick discovery of sentimental contents of movie and hotel reviews conducted with two supervised machine learning algorithms: K-Nearest Neighbour (K-NN) and NB [31]. For movies review, NB gave far better results than K-NN but for hotel reviews these algorithms gave minor, almost similar accuracies. The researchers suggested testing the results with random forest, SVM etc. and trying to implement a new algorithm utilizing the benefits of the both algorithms.

One of the challenges with SA is the large feature space extracted. Many researchers focus towards this area and Fuzzy Rough Set based approaches provide substantial amounts of solutions to overcome them. Chena et al. [32] proposed an approach for drug reviews using fuzzy rough feature reduction. They used Random Forest algorithm with fuzzy-rough QuickReduct feature selection to obtain significantly reduced features and it improved the accuracy. Further they noticed that it does not consume additional run-time overheads and subsequently improve the performance. It is evident that Fuzzy Rough Feature Selection (FRFS) showed promising results and this is a nice start for the future researchers. They believed that this framework may bring tremendous benefits for the real world health sector with cost-efficiency savings to real-world healthcare analysis on large scale data with enormous data. As future development, the overall performance of the framework can be increased while focusing on alternative approaches for learning classifiers to ensure accuracy when dealing with NLP.

A fuzzy rough set-based feature selection algorithm has been used for hierarchical feature selection with sibling strategy and revealed that it is more efficient and more versatile [33]. Further it shows that the classifier resulted with higher performance with establishing the efficiency and effectiveness. It opens new research trends by combining fuzzy rough sets with hierarchical feature selection problems.

Many researchers now combine fuzzy rough set theory with other technologies to improve classifiers. Genetic Search Fuzzy Rough (GSFR) is one of the feature selection algorithm used by [34] using the evolutionary sequential genetic search technique with fuzzy rough set theory to early identification of cancer. This is an extended approach of Fuzzy-rough nearest neighbor (FRNN) classifier. This classifier outperforms with number of features, accuracy, and precision, recall, F-measure and computation time compared to other classifiers. This research opens new doors to hybridize fuzzy rough sets with both Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) in order to improve further.

A new approach based on FRNN proposed by R. Jensen and Cornelis [35] tested over nine data sets. FRNN-FRS uses the traditional operations, a t-norm and an implicator. FRNN-VQRS is the fuzzy quantifier-based approach used. It resulted that, FRNN outperforms both FRNN-O, as well as the traditional fuzzy nearest neighbour (FNN) algorithm. This research opens new investigations by providing explanations of the impact of the choice towards the fuzzy relations, connectives and quantifiers. The accuracy of the classifier upon the feature selection preprocessing is another new research area arising from this.

The above literature review indicates that a sufficient amount of studies have been conducted by various authors during the last decade. It explored that the SA in previous work was to find out whether the expressed opinion in the document or sentence is positive, negative or neutral. Many researchers have studied different techniques for sentiment analysis like NB classifier, SVM algorithm, Rough-Fuzzy classifiers etc. for the SA. NB performed well with most of the cases and easy to implement. SVM obtained favorable results with prediction accuracy while bringing fast evaluation of the learned target function. However, it consumes much computational time, is expensive in memory and requires long training time [36]. The text pre-processing plays a vital role in terms of the accuracy and the performance of the classifier. Interestingly, it has been found that the n-grams used as features in classification improve the result, in some cases the unigrams perform well. Furthermore, the researchers conduct their studies in different domains (product reviews, political sentiment, and news). Decent amount of related prior work has been done in SA. They have contributed to this domain a lot and there are still some gaps to be filled as mentioned in the literature review. Hence sentiment analysis has become a popular field for research work. It is very useful for academic as well as for business purposes.

III. Text Classification Framework with Supervised ML Techniques

Through the knowledge observed with the literature review, an abstraction model of a generic framework for SA in the context of social media depicted in Fig. 2.

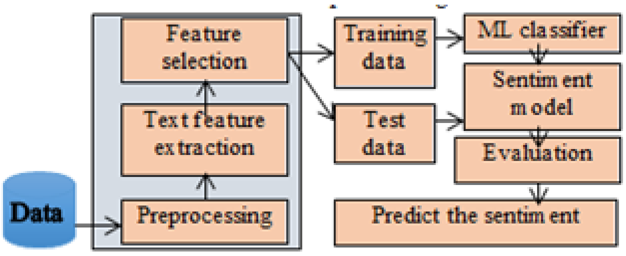

Fig 2 A common Framework for SA - This is a general model for the SA consisting of a number of phases. Both training and testing data should be pre-proceed. Then the feature extraction and selection should be done. The ML classifier developed and trained with

Data preparation plays a vital role since the real data improves the accuracy and the performance. SNW contains data such as comments, tweets which are unstructured representing the opinion of people. Thus, the data preprocessing needed to be done by removing unimportant or disturbing elements such as URLs and hashtags in tweets. Furthermore, remove punctuations, extra blank spaces and vowels repeated in sequence at least three times, convert emoticons into tags, convert text to lower case [37] are necessary follow up states. Stop word removal helps faster processing as well as in dimension reduction in terms of space requirement. Next, commonly used normalization procedures: stemming and lemmatization are to be done [38]. At the final stage the tokenization [39] has to be done in order to split a string into a list of tokens.

Text features are extracted from the above pre-

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (4)$$

processed data. Many researchers use various techniques such as Term Frequency-Inverse Document Frequency (TFIDF) , Bag Of Words (BOW), word embedding, word count, noun count etc [40]. At the end of this step the researchers found that the amount of features extracted is too large to be used directly. Hence, the feature reduction is important at this stage. Chi-square score is a feature reduction technique used by many as the attribute evaluation metric in order to obtain the high information features [41], [42]. Now the Corpus will be split into two data sets, training and test. The training data set will be used to fit the model and the predictions will be performed on the test data set. The data set is now passed into a classification algorithm such as NB, SVM etc. After training the model, the evaluations have to be done for the classifier.

A confusion matrix [43], also known as an error matrix is a special kind of contingency table with two dimensions ("actual" and "predicted"), and identical sets of "classes" in both dimensions that allows visualization of the performance of an algorithm, typically with supervised learning as shown in Table 2 where: P = positive; N = Negative; TP = True Positive; FP = False Positive; TN = True Negative; FN = False Negative.

Table 2 The confusion matrix in abstract terms

		Actual Class	
		P	N
Predicted class	P	TP	FP
	N	FN	TN

From classification point of view TP, FP, TN and FP are used to compare labels of classes [44]. Based on the data of the confusion matrix, following measures are suitable for evaluating performance of classifiers.

Precision measures the exactness.

$$Precision = \frac{TP}{TP + FP}$$

Recall measures the completeness/sensitivity.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

F-measure combines both precision and recall. The F1 score is the harmonic mean of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall).

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

Accuracy is the portion of all true predicted instances against all predicted instances.

Almost all classifiers developed by researchers were evaluated empirically by using above indexes.

Now the classifier is ready for the prediction. The above proposed framework could serve a basis for future works as an extensible and complete guideline for SA.

IV. Challenges and Future Work

There are some weaknesses and limitations within the domain of the sentiment classification identified as obstacles that affect the efficiency and effectiveness of the classifier.

The detection of spam and fake reviews - This is always difficult with human centric corpus.

Duplicates of the same tweet presents.

Word Sense Disambiguation (WSD) - Since this is human centric and opinion based, a word that is considered to be positive in one situation may be considered negative in another situation. Therefore, WSD is a basic and on-going issue that occurs with text mining [45]. Different people express their opinions in different ways. There might be a small difference between the texts, however the meaning can be drastically different from the opinion.

Identifying the entity - This is one of the challenges in opinion mining. A text may have multiple entities associated with it. In the same text it may represent both negative and positive polarity.

The negation problem - This is a problem which can direct to a completely wrong decision as it reverses the polarity of the word.

The intensity of the opinion - This is the degree of polarity which becomes problematic when finding the sentiment.

Sudden changes in the same tweet.

The domain - In some cases, the domain is different and it has hidden meaning for some of the words where context matters.

Demographics bias - This is associated with the comments/tweets and however it was neglected even when it is well known that social media is not a random sample of the population.

Humor and sarcasm - Some people use positive or intensified positive words to express their negative feelings towards an entity. This plays a major role and should take precautions for that.

Comparisons - It is difficult to explore the polarity when the tweets are to compare two entities.

There are many future directions associated with SN.

Expansion for multiple languages - Analyzing sentiment using the only English language (the most common lexicon source is WordNet) isn't worthy of every time since many people use their native language to communicate via SMW and this opens a new research area to create lexica, corpora and dictionaries resources in the context of other natural languages.

Feature selection and reduction plays a vital role in text mining. Feature space reduction approaches while ensuring the performance by removing redundancy be a future direction for research work [46].

The literature reveals that the Fuzzy Rough Sets based approaches effectively improved the accuracy and accomplished higher consistencies. Hence, a research gap arises to increase the overall performance together with above identified measures.

Find new approaches to guarantee accuracy since text mining always deals with NLP.

The recent finding shows the need of fuzzy rough sets based approaches and hierarchical feature selection problems to be addressed to improve the accuracy and the performance of the SA classifiers.

The hybrid concept is much prominent with SA and there are new research areas open for fuzzy rough sets with both Particle Swarm Optimization and Ant Colony Optimization in order to improve classifiers further.

- Experimentation of n-gram models with incorporating the contextual knowledge is rampant and another direction is to work with real live streaming text data.

V. Conclusion

It has been observed that the performance of the classifier depends on the algorithm, and various feature selection schemes. Thus no classifier alone provides complete accuracy and performance whereas need to combine additional features. The use of N-gram feature increases the accuracy of the classifier with improved prediction results as with increased sentiment length. Machine learning approaches surrender best results compared to other approaches in SA. There is a dependency with the domain where a trained classifier on a particular domain does not work accurately with another domain. There are many new research areas open in the field of rough fuzzy SA classification techniques as discussed in this paper such as: increase the accuracy by improving the overall performance, hierarchical feature selection, hybridize of fuzzy rough sets with both PSO and ACO. A common guideline in a structured way is necessary since many researchers actively work in this field and the proposed framework is fulfilling that. As a future work, this framework can be enhanced with new findings by providing the ability for users to plug new components where necessary. The authors wish to test this framework with a case study in the political domain to predict the sentiment towards the candidates in an election by using SMW.

References

[1] Turney, P.D.: Thumbs up or thumbs down?, in Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of ACL-02, 40th Annual Meeting of the Association for Computational Linguistics , Philadelphia, pp. 417-424, July 2002

[2] Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment Classification using Machine Learning Techniques. In: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), pp. 79-86, July 2002, doi: 10.3115/1118693.1118704

[3] Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: A survey. In: Ain Shams Engineering Journal, vol 5, no. 4, December 2014

[4] Payne, J.G.: The Bradley effect: Mediated reality of race and politics in the 2008 US presi-dentialelection. In: American Behavioral Scientist, vol. 54, no. 4, pp. 417–435, 2010

[5] Ceron, A., Curini, L., Iacus S.M., Porro, G.: Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. In: New Media and Society, vol. 16, no. 2, pp. 340–358, 2014

[6] Meghashree, K., Radhika, S., Shilpashree, A., Dinni, S., Monika, P.: Survey Paper on Algorithms used for Sentiment Analysis. In: International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 8, no. V, May 2020

[7] Devika, M.D., Sunitha C., Ganesha, A.: Sentiment Analysis:A Comparative Study On Different Approaches. In: Fourth International Conference on Recent Trends in

Computer Science & Engineering, Procedia Computer Science 87 (2016), pp. 44 – 49, 2016, doi: 10.1016/j.procs.2016.05.124
[8] Srinivas, K., Rao, G. R., Govardhan, A.: Rough-Fuzzy Classifier: A System to Predict the Heart Disease by Blending Two Different Set Theories. In: Arabian Journal for Science and Engineering, vol. 39, no. 4, pp. 2857–2868, 9 February 2014, doi: 10.1007/s13369-013-0934-1
[9] Keerthika, U., Sethukkarasi, R., Kannan, A.: A ROUGH SET BASED FUZZY INFERENCE SYSTEM FOR MINING TEMPORAL MEDICAL DATABASES. In: International Journal on Soft Computing (IJSC), vol.3, no.3, August 2012 doi: 10.5121/ijsc.2012.3304
[10] Srividya, K., Sowjanya, A.M.: Sentiment analysis of facebook data using naïve Bayes classifier. In: International Journal of Computer Science and Information Security (IJCSIS), vol. 15, no. 1, January 2017
[11] Beck, M.S.L.: Election forecasting: principles and practice. In: The British Journal of Politics and International Relations, vol. 7, no. 2, pp. 145-164, 2005
[12] Fumagalli, L., Sala, E.: The total survey error paradigm and pre-election polls: The case of the 2006 Italian general elections. In: ISER Working Paper Series, 2011
[13] Salunkhe, P., Deshmukh, S.: Twitter Based Election Prediction and Analysis. In:International Research Journal of Engineering and Technology (IRJET), vol. 04, no. 10, October 2017
[14] Wong, F.M.F., Tan, C.W., Sen, S., Chiang, M.: Quantifying Political Leaning from Tweets, Retweets, and Retweeters. In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, 2013
[15] Vadivukarassi, M., Puviarasan, N., Aruna, P.: Sentimental Analysis of Tweets Using Naive Bayes Algorithm. In: World Applied Sciences Journal,India, vol. 35, no. 1, pp. 54-59, 2017, doi:10.5829/idosi.wasj.2017.54.59
[16] Prabhu, B. P. A., Ashwini, B. P., Khan, T.A., Das, A.: Predicting Election Result with Sentimental Analysis Using Twitter Data for Candidate Selection. In: Innovations in Computer Science and Engineering, pp. 49-55, 19 June 2019 doi: 10.1007/978-981-13-7082-3_7
[17] Pak, A., Paroubek, P.: Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In: Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, Valletta, Malta, European Language Resources Association 2010, pp. 17-23 May 2010, ISBN 2-9517408-6-7, 1994
[18] Schmid, H.: Probabilistic part-of-speech Tagging Using Decision Tree. In: International Conference on New Methods in Language Processing, Manchester, UK
[19] Shannon, C.E., Weaver, W.: A Mathematical Theory of Communication. In: The University of Illinois Press, Champaign, IL, USA. 1963
[20] Rao, C.S., Prasad, G.S., Rao, V.V.: Prediction and Analysis of Sentiments on Twitter Data using Machine Learning Approach. In: International Journal of Computer Science and Information Security (IJCSIS), vol. 16, no. 8, August 2018
[21] Wei, W., Gulla, J.: Sentiment Learning on Product Reviews via Sentiment Ontology Tree. In: 48th Annual Meeting of the Association for Computational Linguistics. [online] Sweden: Association for Computational Linguistics, pp. 404–413, 2019
[22] Pang, B., Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In: Proceedings of the ACL, 2004
[23] Zainuddin, N., Selamat, A.: Sentiment Analysis Using Support Vector Machine. In: International Conference on Computer, Communication, and Control Technology (I4CT 2014), September 2 - 4, 2014 - Langkawi, Kedah, Malaysia
[24] Bhumika, M.J., Vimalkumar, B.V.: Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis. In: International Journal of Computer Applications (0975 – 8887), vol. 146, no.13, July 2016,
[25] Ahmad, M., Aftab, S.: Analyzing the Performance of SVM for Polarity Detection with Different Datasets. In: Int. J. Mod. Educ. Comput. Sci., vol. 9, no. 10, pp. 29–36, 2017.
[26] Tarlekar, A., Kodmelwar, M.K.: Sentiment Analysis of Twitter Data from Political Domain Using Machine Learning Techniques. In: International Journal of Innovative Research in Computer and Communication Engineering, vol. 3, no. 6, 2015, doi: 10.15680/ijircce.2015.0306084
[27] Elghazaly, T., Mahmoud, A., Hefny, H.A.: Political Sentiment Analysis Using Twitter Data. In: ICC '16: Proceedings of the International Conference on Internet of things and Cloud Computing, 2016, doi: 10.1145/2896387.2896396
[28] Ringsquandl, M., Petkovic, D.: Analyzing Political Sentiment on Twitter. In: 2013 AAAI Spring Symposium, 2013
[29] Kassraie, P., Modirshanechi, A., Aghajan, H.K., Election Vote Share Prediction using a Sentiment-based Fusion of Twitter Data with Google Trends and Online Polls. In: Proceedings of the 6th International Conference on Data Science, Technology and Applications, pp: 363-370, 2017
[30] Raghuwanshi, A.S., Pawar, S.K.: Polarity Classification of Twitter Data using Sentiment Analysis. In: International Journal on Recent and Innovation Trends in Computing and Communication, vol. 5, no. 6, pp: 434 – 439, 2015
[31] Dey, L., Chakraborty, S., Biswas, A., Bose, B., Tiwari, S.: Sentiment Analysis of Review Datasets Using Naïve, Bayes‘ and K-NN Classifier. In: International Journal of Information Engineering and Electronic Business, 2016, doi: 10.5815/ijieeb.2016.04.07
[32] Chen,, Su, T. P., Shang, C., Hill, R., Zhang, H., Shen, Q.: Sentiment Classification of Drug Reviews Using Fuzzy-rough Feature Selection. In: 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2019. Available: 10.1109/fuzz-ieee.2019.8858916 [Accessed 11 May 2020].
[33] Zhao, H. , Wang, P., Hu Q., Zhu, P.: Fuzzy Rough Set Based Feature Selection for Large-Scale Hierarchical Classification. In: IEEE Transactions on Fuzzy Systems, vol. 27, no. 10, pp. 1891-1903, 2019. Available: 10.1109/tfuzz.2019.2892349 [Accessed 11 May 2020].
[34] Meenachi, L., Ramakrishnan, S.: Evolutionary sequential genetic search technique-based cancer classification using fuzzy rough nearest neighbour classifier. In: Healthcare Technology Letters, vol. 5, no. 4, pp. 130-135, 2018. Available: 10.1049/htl.2018.5041 [Accessed 11 May 2020]
[35] Jensen, R., Cornelis, C.: A New Approach to Fuzzy-Rough Nearest Neighbour Classification. In: Chan CC., Grzymala-Busse J.W., Ziarko W.P. (eds) Rough Sets and Current Trends in Computing. RSCTC 2008. Lecture Notes in Computer Science, vol 5306. Springer, Berlin, Heidelberg, 2008
[36] Patil, P., Yalagi, P.: Sentiment Analysis Levels and Techniques: A Survey. In: International Journal of Innovations in Engineering and Technology (IJIET), vol. 6, no 4, April 2016
[37] Angiani, G., Ferrari, L., Fontanini, T., Fornacciari, P., Iotti, E., Magliani, F., Manicardi, S. : A Comparison between Preprocessing Techniques for Sentiment Analysis in

Twitter. In: 2nd International Workshop on Knowledge Discovery on the Web-KDWEB 2016, 2016, Cagliari, Italy

[38] Batanovic, V., Nikolic, B.: Sentiment Classification of Documents in Serbian: The Effects of Morphological Normalization and Word Embeddings. In: Telfor Journal, Vol. 9, No. 2, 2017

[39] Kusrini, Mashuri, M.: Sentiment Analysis In Twitter Using Lexicon Based and Polarity Multiplication. In: 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), pp. 365-368, Yogyakarta, Indonesia, 2019 doi: 10.1109/ICAIIT.2019.8834477.

[40] Ahujaa, R., Chuga, A., Kohlia, S., Guptaa, S., Ahuja, P.: International Conference on Pervasive Computing Advances and ApplicaThe Impact of Features Extraction on the Sentiment Analysis. In: International Conference on Pervasive Computing Advances and Applications – PerCAA 2019, 2019, doi: 10.1016/j.procs.2019.05.008

[41] Kouloumpis, E., Wilson, T., Moore, J.: Twitter sentiment analysis: The good the bad and the omg!. In: Icwsm, vol. 11, pp. 538–541, 2011.

[42] Yang, A., Zhang, J., Pan, L., Xiang, Y.: Enhanced Twitter Sentiment Analysis by UsingFeature Selection and Combination. In: International Symposium on Security and Privacy in Social Networks and Big Data, 2015, doi: 10.1109/SocialSec2015.9

[43] Tripathy, A., Rath, S.K.: Classification of Sentiment of Reviews using Supervised Machine Learning Techniques. In: International Journal of Rough Sets and Data Analysis, vol. 4, no. 1, January-March 2017

[44] Mouthami, K., Devi, K.N., Bhaskaran, V.M.: Sentiment analysis and classification based on textual reviews. In: Proceedings of theInternational Conference on Information Communication and Embedded Systems (ICICES), pp. 271-276, 2013, doi:10.1109/ICICES.2013.6508366

[45] Wang, Y., Wang, M., Fujita, H.: Word Sense Disambiguation: A comprehensive knowledge exploitation framework. In: Knowledge-Based Systems (2019) 105030

[46] Asghar, M.Z., Khan, A., Ahmad, S., Kundi, F.M.: A Review of Feature Extraction in Sentiment Analysis. In: Journal of Basic and Applied Scientific Research, February 18 2014